

NASA			
<b>Report Documentation Page</b>			
1. Report No.		2. Government Accession No.	
3. Recipient's Catalog No.			
4. Title and Subtitle		5. Report Date	
Expert Seeker: a People-Finder Knowledge Management System		6. Performing Organization Code	
7. Author(s)		8. Performing Organization Report No.	
Irma Becerra-Fernandez		10. Work Unit No.	
9. Performing Organization Name and Address		11. Contract or Grant No. NAS5-32337	
Florida International University		USRA subcontract No. 5555-97-74	
University Park Campus			
Miami, Florida 33199		13. Type of Report and Period Covered Final	
12. Sponsoring Agency Name and Address		January 2000 - September 2000	
National Aeronautics and Space Administration		14. Sponsoring Agency Code	
Washington, DC 20546-0001			
NASA Goddard Space Flight Center			
Greenbelt, MD 20771			
15. Supplementary Notes			
This work was performed under a subcontract issued by			
Universities Space Research Association			
10227 Wincopin Circle, Suite 212			
Columbia, MD 21044			
Task 97			
16. Abstract			
<p>The first objective for this report was to perform a comprehensive research of industry models currently being used for similar purposes, in order to provide the Center with ideas of what is being done in area by private companies and government agencies. The second objective was to evaluate the use of taxonomies or ontologies to describe and catalog the areas of expertise at GSFC. The creation of a knowledge taxonomy is necessary for information extraction in order for The Expert Seeker to adequately search and find experts in a particular area of expertise.</p> <p>The requirements to develop a taxonomy are: provide minimal descriptive text; have the appropriate level of abstraction; facilitate browsing; ease of use and speed of data entry are critical for success; customized to the organization and its culture; extent of knowledge areas; expandable, so new skills could be develop; could be complemented with free text fields to allow users the option to describe their knowledge in detail.</p>			
17. Key Words (Suggested by Author(s))		18. Distribution Statement	
taxonomies		Unclassified--Unlimited	
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	1	



**Goddard Space Flight Center**  
**Expert Seeker: a People-Finder Knowledge**  
**Management System**

**Final Report**

**Grant Number NAS5-32337**

**March 1st 2000 - September 26th 2000**

**Irma Becerra-Fernandez Ph.D.**  
**Florida International University**



## Acknowledgements

The Author wishes to acknowledge the following organizations and individuals for their contributions to this project:

NASA/Goddard Space Flight Center (GSFC)-Center for Excellence in Space Data and Information Systems (CESDIS) for financial support to this project, entitled **"Expert Seeker: a People-Finder Knowledge Management System"** Grant Number NAS5-32337.

Dr. Milton Halem, Dr. Susan Hoban, Nancy Laubenthal, Nora Farrell, Steve Naus, Robert Peirce, Jerome Bennett, Tim Burch, Pravin Gohel, Dr. Daniel Russ, Joel Sachs, Sue Kaltenbauch, and Mike McGunigale.

The students at Florida International University: Claudia Alarcon, Gabriela Rizzo, Greg Copeland, Haniel Pulido, Hector Hartmann, Hernan Santiesteban, Jaclyn Don, Juan Rodriguez, Lusally Mui, Lynn Hernandez, Maria Ray, Nelson Ospina, Rigoberto Fernandez and Thomas Pla.

## TABLE OF CONTENTS

I.	Introduction.....	5
II.	Objectives Accomplished.....	5
III.	Development Meetings .....	6
	A. Kick-off at GSFC .....	6
	B. The First Expert Seeker Prototype.....	7
IV.	First Usability Test of the First Expert Seeker Prototype: Feedback and Proposed Enhancements.....	8
V.	The Second Expert Seeker Prototype.....	11
	A. Research Results and Changes Completed to Date .....	11
	B. Web Text Mining Component.....	12
VI.	Using Expert Seeker.....	13
	A. Home .....	14
	B. Search by Name.....	15
	C. Search by Expertise .....	15
	D. Directorate Search .....	16
	E. Advanced Search .....	16
	F. Web Page Search .....	18
VII.	Operators.....	18
	A. Concept Operators.....	19
	B. Evidence Operators.....	19
	C. Proximity Operators .....	19
VIII.	Challenges.....	20
IX.	Immediate Objectives.....	21
X.	The Technologies used to develop Expert Seeker.....	21
XI.	Relevant publications related to this project .....	21
XII.	Attachments .....	22

## TABLE OF FIGURES

Figure 1: Prototype I Expert Seeker Interface.....	7
Figure 2: Redesigned Graphical User Interface - Directorate Search .....	9
Figure 3: Expert Seeker's Redesigned GUI Main page .....	11
Figure 4: Expert Name.....	15
Figure 5: Expertise.....	16
Figure 6: Directorate .....	17
Figure 7: Advanced Search .....	17
Figure 8: Web Page Search.....	18

## **I. Introduction**

Initial development of Expert Seeker was originally funded by NASA Kennedy Space Center (KSC), through the Faculty Awards for Research (FAR-99) grant. Expert Seeker is being developed at the Knowledge Management (KM) Laboratory at Florida International University (FIU), under the direction of Dr. Irma Becerra-Fernandez, Principal Investigator. This grant by the Goddard Space Flight Center (GSFC)-Center for Excellence in Space Data and Information Systems (CESDIS), seeks to synergize with the efforts at KSC to create a web-based application, which will serve as a centralized repository of experts within NASA. The original development of Expert Seeker followed from the recommendations from the Knowledge Management Assessment performed at KSC in 1998, which affirmed the need for a Center-wide repository that would provide KSC employees with Intranet access to experts with specific backgrounds.

The benefits of Expert Seeker are:

- Assist NASA GSFC employees in identifying team members who have the appropriate skills necessary to staff projects, and how to contact them in order to determine if those experts are available.
- Assist NASA GSFC management in performing an intellectual capital gap analysis, to identify areas of expertise where investments should be made to encourage employees to further develop their skills.
- Minimize the loss of knowledge and expertise of employees that leave or retire from NASA GSFC.
- Assist NASA GSFC employees in identifying opportunities for innovation and minimizing duplication efforts.
- Provides NASA GSFC employees access to cutting-edge knowledge management systems (KMS) that are aligned with NASA's business strategy.
- Provide external organizations with a KMS that can be used to effectively identify a point-of-contact knowledgeable in an area of competency specific to NASA.
- Help create an environment of partnership and rapid development that is required in today's business economy, where there is an increasing need to pull together teams of experts who may work for NASA, Universities, research laboratories, or other related industries for collaboration in order to quickly solve problems.
- Provide NASA GSFC experts more outside exposure and publicity.
- Store in a centralized repository the various competencies of the employees of NASA GSFC, including items such as past projects and awards, which are not typically captured by most Human Resources applications.
- Unify the countless data collections into a web-enabled repository that could easily be searched for relevant data.

## **II. Objectives Accomplished**

Our first objective prior to the development of Expert Seeker was to perform a comprehensive research of industry models currently being used for similar purposes, in order to provide the Center with ideas of what is being done in this area by private companies and government agencies. We also proposed the creation of an Advisory Group that would guide the development team according to the requirements of NASA GSFC.

Another preliminary task was to evaluate the use of taxonomies or ontologies to describe and catalog the areas of expertise at GSFC. The creation of a knowledge taxonomy is necessary for information extraction in order for Expert Seeker to adequately search and find experts in a particular area of expertise. Standard taxonomies were also studied; including those published by the Department of Labor and the Library of Congress. Ultimately a set of skills and sub-skills from GSFC's Manpower Assessment Reporting System (MARS) database was integrated with Expert Seeker.

The requirements to develop a taxonomy are:

- Provide minimal descriptive text.
- Have the appropriate level of abstraction:
  - Too low = too complicated to use.
  - Too high = insufficiently describe the knowledge areas.
- Facilitate browsing.
- Ease of use and speed of data entry are critical for success.
- Customized to the organization and its culture.
- Extent of knowledge areas.
- Expandable, so new skills could be developed.
- Could be complemented with free text fields to allow users the option to describe their knowledge in detail.

### **III. Development Meetings**

#### **A. Kick-off at GSFC**

Dr. Irma Becerra-Fernandez, Principal Investigator, and Hector Hartmann, research associate, attended the Knowledge Management kick-off Workshop on February 22, 2000 at GSFC. During this meeting, Dr. Becerra-Fernandez made a presentation of a preliminary prototype of Expert Seeker. She also delivered a brief commentary on KMS, the purpose of these systems, and their benefits. Her presentation addressed research performed to date on systems similar to the proposed Expert Seeker, and discussed the schedule for completion and delivery of this project to GSFC.

Following this introductory meeting, preliminary contacts were made with Jerome Bennett to request access to GSFC's Intranet and GSFC's X.500. Students from the KM Lab also contacted Robert Peirce for assistance in gaining access to the MARS database. Finally, Nancy Laubenthal provided a set of hyperlinks to Goddard Space Flight Center's

web pages for the general public (the *stars* domain), which served as an initial research platform for the text mining portion of this project.

## B. The First Expert Seeker Prototype

The Expert Seeker initial prototype that met the user specifications set forth by the GSFC Advisory Group was demonstrated during the second progress meeting, held in Miami, Florida on June 1, 2000. This first prototype of Expert Seeker implemented the use of career summaries to help users locate expertise from short biographies submitted by the experts themselves. This information will be important in particular for contractors, since little descriptive data exists for them in the NASA databases. Expert Seeker allows the user to search for specialists by name, field of expertise and directorate (Figure 1).

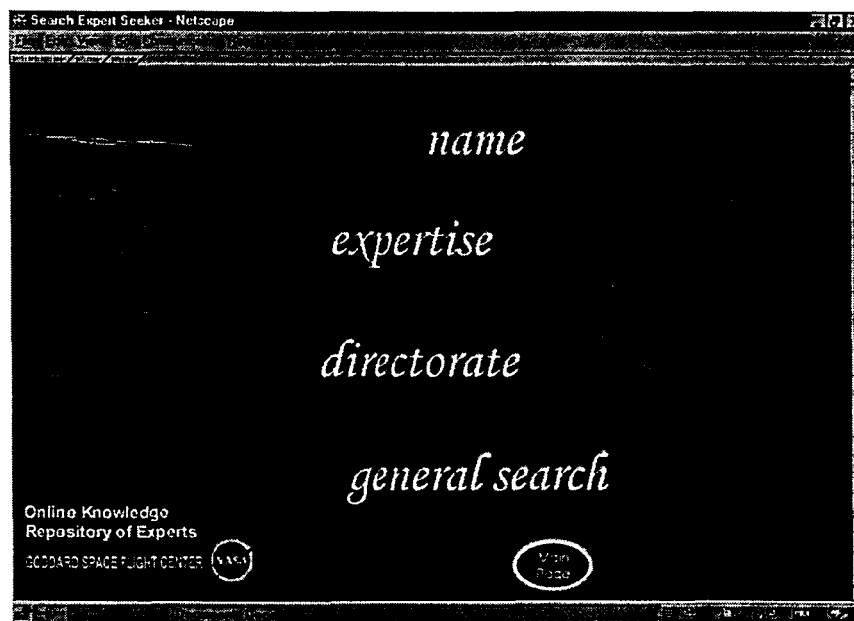


Figure 1: Prototype I Expert Seeker Interface

Selecting the "*Name*" icon while browsing through Expert Seeker accesses "*Name Search*". This option offers the user the ability to search for experts at any directorate within GSFC. The user encounters a "drop-down menu" where they have the option to scroll through the names or type the name in the field. The names of the experts in Expert Seeker come from Goddard's X.500 database, which provides complete contact information. However, the first version of the X.500 database only contained names, phone numbers and organization codes for some of the GSFC employees.

Similarly, electing the "*Expertise*" icon while browsing through Expert Seeker accesses "*Expertise Search*", which offers the user the ability to search for experts depending on their areas of specialization. Utilizing the "drop-down menu" helps the user select a general skill and a second "drop-down-menu" helps the user select a sub-skill or



“area of specialization”. The set of skills and sub-skills, were provided by Robert Peirce and came directly from the MARS Database.

The "*Directorate Search*" offers the user the ability to search for experts depending on their department. The first "drop-down menu" enables the user select a directorate and the second "drop-down menu" gives the user the option to further narrow the search to a specific branch or office. The directorate database was assembled at the KM Lab after browsing through all GSFC's directorates web pages.

Selecting the "*General Search*" icon while browsing through Expert Seeker accesses "*General Search*", which allows the user the ability to search for experts using keywords and as well as their career summaries.

#### **IV. First Usability Test of the First Expert Seeker Prototype: Feedback and Proposed Enhancements**

Steve Naus performed a usability analysis of the first version of Expert Seeker, which provided us with the feedback necessary to further enhance the system. The following suggestions made by Steve Naus were implemented in order to make the system more user friendly.

- “One overriding problem is that it takes too many "clicks" to access the information. Some examples: On the search pages, first you choose the type of search, then the category, then sometimes a sub-category.”

For the "*Directorate Search*," the Expert Seeker Prototype I forced the user to select a "Directorate", then to press a submit button, only to find another menu from which to select a branch or office within the Directorate. It took several "clicks" for the user to submit the final search. We have optimized the Directorate Search by redesigning Expert Seeker's graphic user interface and by unifying the "Directorate Menu" and the "Office Menu" into one page, thus eliminating one "click"(Fig. 2).

Furthermore, we have modified the code to automatically display the "Offices" related to a particular "Directorate." In other words, when a user chooses a "Directorate" from the drop-down-menu, the "Office Menu" will automatically display the appropriate sub-directory. The user will have the option of searching within the whole directorate or narrowing down the search to a particular office. The database that defines the hierarchy of Directorates and Offices was developed in-house at the KM Lab. Similar modifications were also done in the "*Expertise Search*" that uses the knowledge taxonomy (skills and sub-skills). The skills and sub-skills information were taken directly from the MARS database, following the original format of that database.

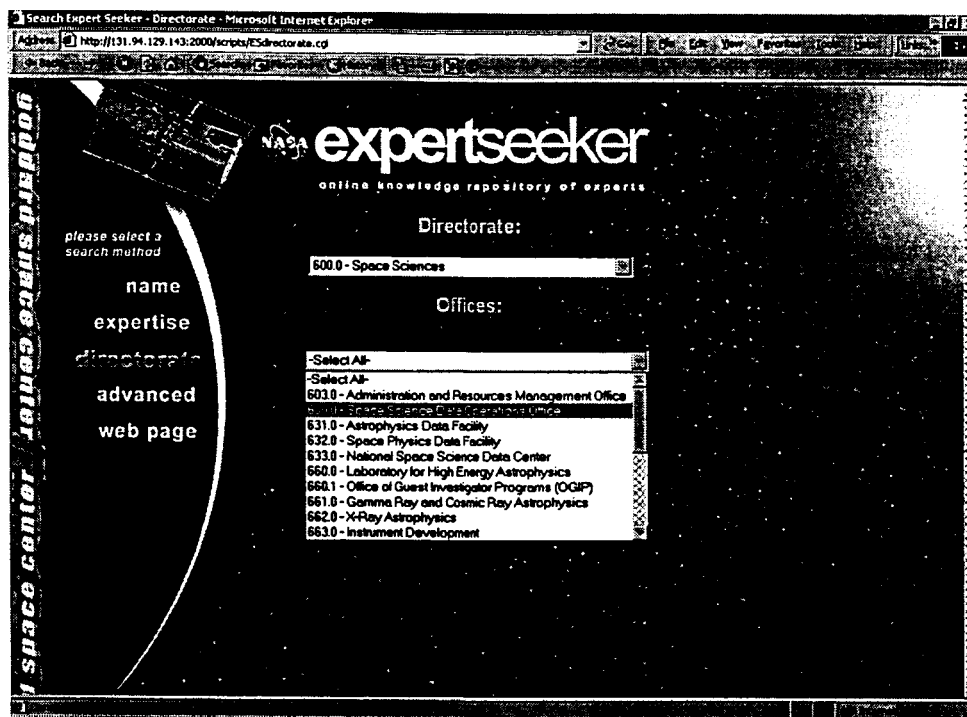


Figure 2: Redesigned Graphical User Interface - Directorate Search

- "It would be useful to allow combined searches on more than one category."

When the system was presented to the advisory group, we included a ***General Search*** that was not yet fully active. This search mode was included to provide the user with the ability to search more than one category. For the final version of Expert Seeker, the "Advanced Search" mode can search for experts by combining various search modes such as name, directorate, and career summaries, further refining the search.

- "The "experts" information page: Here you have some summary information and then have to click to get a "career summary" and "intellectual capital". It would be better to go ahead and display that information on the page - there is plenty of room."

The experts' information page was originally designed to include e-mail, fax numbers, room, building, achievements, past and present projects, awards, education, etc. The first version of the X.500 database directory we received from GSFC only contained the name, organization code, and phone numbers of the experts. A few weeks following our meeting in Miami, Robert Peirce sent us a version of the X.500 database that contained fax number, emails, building number, and room number for each expert. This information has been integrated into the database and is displayed in the new version of Expert Seeker.

More recently, Robert Peirce sent us a DIREX database from the Human Resources department. Several fields from this database included information such as awards, education, and training, etc. This data has already been integrated into Expert Seeker, and is now displayed on the results page.

Steve Naus also mentioned that there is no single source for the "Intellectual Capital" or "Document Repositories". According to him, it is a "data-mining nightmare" that they are just beginning to investigate. He seriously doubted that there would be anything available for this effort. Therefore, we eliminated this section and instead created a link to display training and awards information.

- "The font is too small on some pages. Also don't use italics for the career summary - make it far too difficult to read. The font used for the links on the main page is very fuzzy - sharpen it up"

We have changed the font type and size throughout the site to fix this problem. All text in the new version of Expert Seeker is now easier to read and understand.

- "The graphic changes to black with a "sunspot" once you enter the site and the links move to the opposite side of the circle. Keep navigation consistent on all of the pages."

The first Expert Seeker prototype delivered for GSFC presented a main page with three options:

1. Information about Expert Seeker, News, FAQ's, and a contact.
2. Career Summary Upload.
3. Enter the System, and begin a Search.

We completely re-designed the GUI and created a new main page (Fig. 3) which offers several advantages. The first is that the GUI includes a link not only to the first two search alternatives, but it also offers all the options from the search menu. Another advantage of the new GUI is that these navigation buttons do not change position as the user accesses other pages. They are displayed at all times, and as a result, the user is able to switch search modes without having to go back to the main page, allowing for easier and more consistent navigation.

- "The graphic on the main page does not fill the whole browser window. The white circle is cut off at the edge - makes me wonder if there should be something there"

The previous GUI was designed to be viewed at a resolution of 800X600 pixels, therefore viewing it using 1024x768 would result in a blank area to the right of the screen. Given this limitation, we redesigned the GUI in order to create an interface that could be viewed effectively, using any of the above resolutions.

- “The links turn white when the cursor is moved over them. This is opposite of normal - link usually get darker during mouse-overs...The links once you enter the site don't high-light - the consistency thing again.”

This issue was also resolved with the development of the new GUI. Links now turn darker during mouse-overs. In the new design of Expert Seeker, the navigation buttons are integrated with the main page, making every aspect more consistent and therefore easier to use.

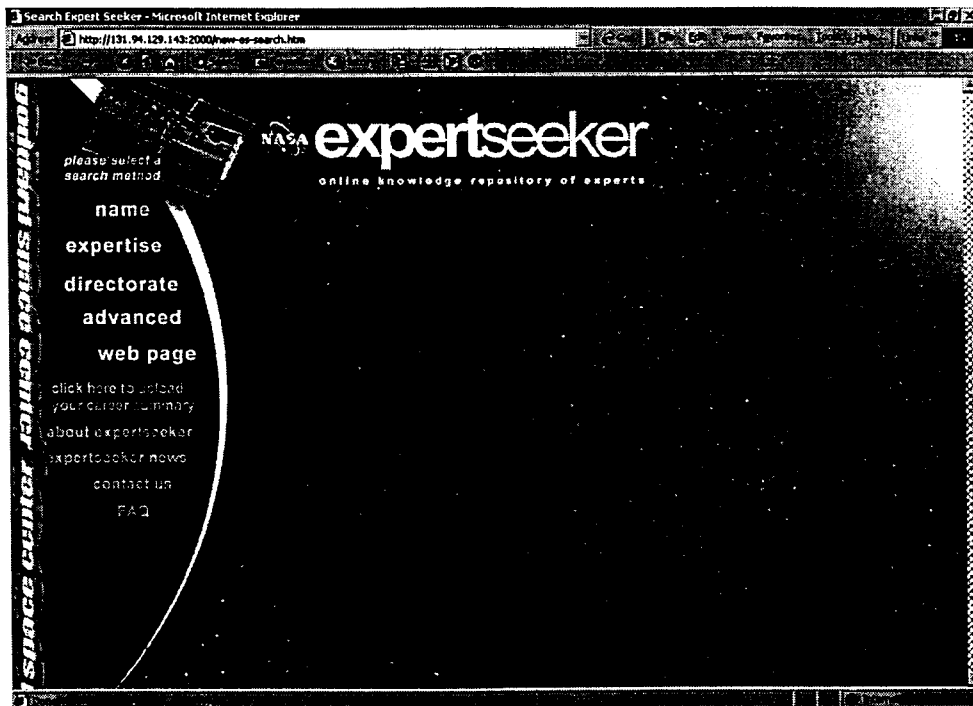


Figure 3: Expert Seeker's Redesigned GUI Main page

## V. The Second Expert Seeker Prototype

### A. Research Results and Changes Completed to Date

The second version of Expert Seeker can viewed at <http://131.94.129.143:2000>. This new version was made available to introduce changes to the initial prototype presented on June 01, 2000. The changes in the second version primarily involve the development of a new and improved GUI, the integration of the Web-based text mining functionality, and integration of the three methods of expertise search into one results page. Furthermore, drop-down menu boxes were eliminated to allow more natural, free text entry into the Name, Directorate and Expertise search options. Drop-down menu boxes are currently used only in the taxonomy-based expertise search. Furthermore, the “*Expertise Search*” mode also allows searching for expertise that combines more than one search category: web-based text mining of GSFC Web-Sites, database extraction

using the MARS Knowledge Taxonomy; or self-assessment text mining using the career summaries. The results of the three launched searches are integrated into one results page.

Dr. Susan Hoban pointed out that while Expert Seeker's design allowed users to upload their career summaries, it did not offer an option to edit or modify them. We modified the code to allow users to edit their own summaries, and limited the field to 1600 characters or less, thereby requiring employees to prioritize salient points of their career in a concise manner.

## **B. Web Text Mining Component**

The text mining portion of Expert Seeker is based on a traditional Information Retrieval (IR) techniques with some additional features. An IR system typically consists of an inverted file, which is a sequence of words that reference the group of documents the words appear in. These words are chosen according to a selection algorithm that determines which words in the document are good index terms. In a traditional IR system, the user enters a query, and the system retrieves all documents that match that keyword entry. The IR technique used for this Expert Seeker goes one step further. Since the user is looking for experts in a specific subject area, instead of documents, the system determines who the experts are according to proper names that appear in the documents (excluding webmasters and curators) that match the keyword, and returns the names of those NASA-Goddard employees. Basically, in addition to keyword selection, the indexing process determines employee name information related to each document and indexes these accordingly. When a user enters a query, all relevant documents are retrieved from the database. The employee names that are associated with those documents are extracted. The system then calculates a score for each name according to the number of documents returned, and ranks each employee accordingly. The employee information is then displayed to the user.

The indexing process was carried out in four stages. First, all the relevant data was transferred to a local directory for further processing. In this case, the data included all the web pages in the Goddard Space Flight Center domain. This was done with a simple web-mirroring tool walled *WGet*.

The second stage was to programatically examine each HTML file and identify all instances of Goddard employee names. This was done using the name data from the X.500 personnel directory databases, which were provided by Goddard. Each name entry is referenced by last name. All employees with the same last name were placed in the same row. Furthermore, each name was stored in a database in all possible variations, for example: John A. Smith was stored as J.A. Smith; J. Smith; Smith, J.A.; and Smith, J. The name finder first determines what last names are present in the web document and then identifies which full name matches for each name type referenced by that last name.

The third stage involves identifying keywords within the HTML content. This is done using a word frequency calculation. First the text is broken up into individual words. This is done through regular expression matching. Any sequence of alphabetical characters is recognized as a word while punctuation, numbers, and whitespace characters are ignored. The resulting list of words is processed to determine if a word was included in a stoplist. (A stoplist is a group of words that are not considered to have any indexing value. These include common words such as "and", "the", and "there".) The resulting list of words was then processed with a stemming algorithm. A stemmer is used to remove the suffix of a word. This is done to group together words that may be spelled differently but have the same semantic meaning. A person who types "astronomical" as a query term would most likely also be interested in documents that match the term "astronomy".

Once the stemming process is completed, the fourth stage involves calculating the frequency of each term. The thirty most frequently occurring words in each document are then chosen as index terms. In Information Retrieval, William Frakes and Ricardo Baeza-Yates detail an algorithm for selecting index terms (Selection by Discriminant Value). This involves calculating the average similarity for each document in the collection, both with and without a potential index term. A positive value means that a potential index term decreases the average similarity by its presence and thus is a good discriminator for the documents within the collection. This method is employed within the indexing process. It was determined that the thirty most frequently occurring words consistently scored positively as discriminators and hence were good index terms. Since the maximum threshold for number of index terms per document was thirty, this method was used for keyword selection.

These keywords have a twofold purpose. First, they are used to quickly associate employees with recurring skill terms. It is assumed that if an employee is continually mentioned in documents that have similar associated keywords, then that person has knowledge about some or all of these keywords. These keywords can also be used in future work for clustering similar documents into topic areas. Further work includes taxonomy construction from these keywords and the development of a query relevance feedback system that suggests query terms that are related to the query entered by the user.

The text-mining software component was not integrated to the first prototype of Expert Seeker. Major changes in the code had to be implemented to include the Web-based search mode with the "*Expertise Search*" method. "*Expertise Search*" now includes the taxonomy search, the career summary search, and the web-based text mining search. Moreover, changes were also made to display the results from the three different search methods simultaneously.

## VI. Using Expert Seeker

This section contains an abridged User Manual, in order to provide an overview of the various options offered by this Expert Seeker prototype.

After accessing this first page the user has the option to search for specialists using various criteria. Clicking on any of the topics accesses each search mode. This allows the user to go onto the next page of his/her choice. In addition, these criteria are displayed throughout Expert Seeker, below the main heading of each search page, allowing the user to browse easily and efficiently through the system:

- A. *"Home"*: returns to the home page.
- B. *"Name"*: accesses the Name Search functionality.
- C. *"Expertise"*: accesses the Expertise Search functionality.
- D. *"Directorate"*: accesses the Directorate Search functionality.
- E. *"Advanced"*: accesses the Advanced Search functionality.
- F. *"Web Page"*: accesses the Web Page Search functionality.

The results of each search are displayed in groups of fifteen. If there are more than fifteen results, the user can click on the link entitled *"Next Results"* located at the bottom of the page. When the last page of results is displayed, the words *"Last Results"* indicate there are no further pages.

At the bottom of each page of results, there are options for more information about each expert:

- *"Career Summary"*: synopsis of the employees' career. The Career Summary will be uploaded from the main page by the expert.
- *"Training"*: displays a list of all the training courses the employee has taken and will even include the dates when these training sessions were attended.
- *"Awards"*: Displays all the employee's awards in the order they were received.
- *"Honors"* link will display a list of honors received by the employee, again, in chronological order.

#### A. Home

The main page of Expert Seeker is the springboard for beginning any search within this system. In addition to the five expert search options, it also offers an information menu.

The employees have the ability to upload their career summaries from this page, by clicking on the icon aptly titled *"click here to upload your career summary."*

The last four information options are currently under construction, including:

- "About Expert Seeker"
- "Expert Seeker News"
- "Contact Us"
- "FAQ"

## B. Search by Name

Selecting "*Name*" while browsing through Expert Seeker accesses "*Expert Name*". This option offers the user the ability to search for experts at any directorate within GSFC by typing the name or part of the name in the space provided. The search can be executed by typing in names, last names, or just the first letter(s) of the first or last name. If searching for more than one expert the user has to separate the names with comas (Figure 4).

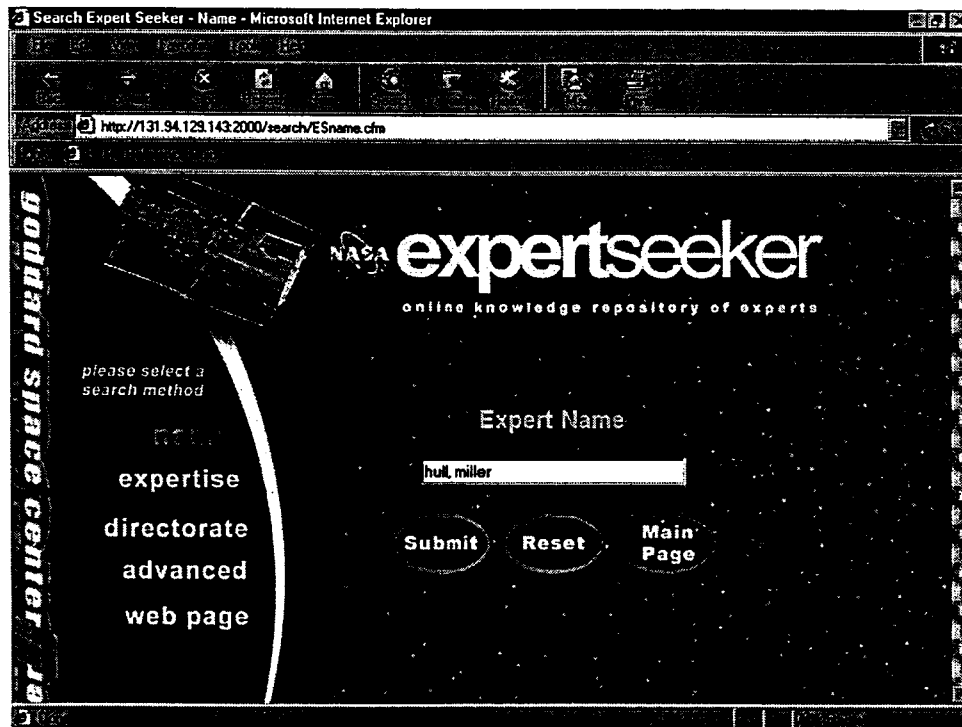


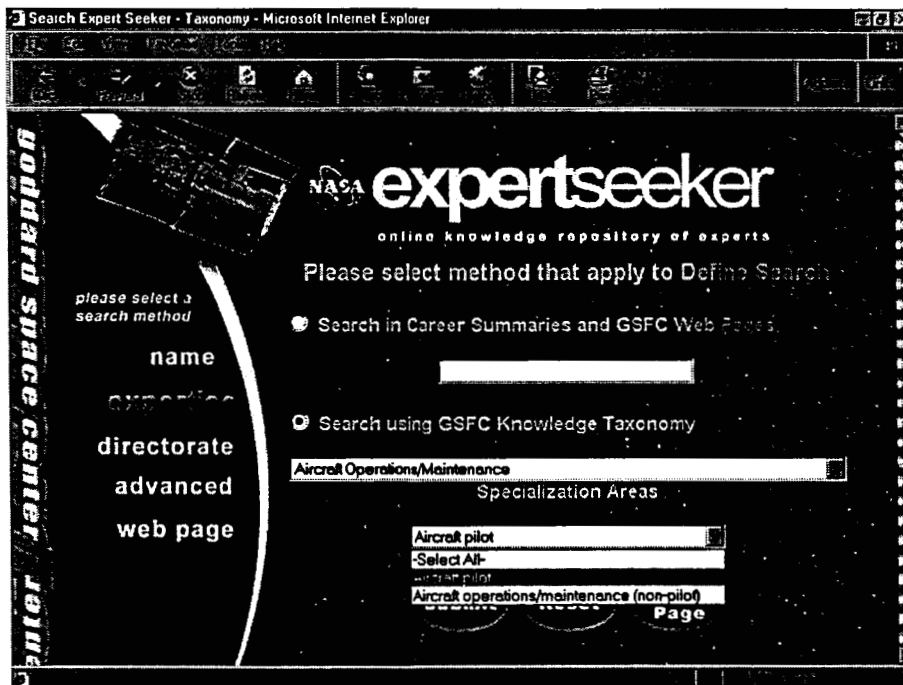
Figure 4: Expert Name

## C. Search by Expertise

This search offers the user the ability to find experts in one of two ways. The user will first be required to select the type of search desired. If the first option is chosen, the user enters a specific keyword that describes the desired field of expertise. Expert Seeker will look for that keyword in the experts' "*Career Summaries*" and web pages, and display the names of any matching results.



If the second option is selected, the user can search for experts depending on their expertise field and area of specialization from the GSFC knowledge taxonomy. The drop-down menu boxes offer a choice of all the different expertise fields applicable GSFC. The first drop-down menu allows the selection of an expertise field, and the second drop-down menu displays the specific sub-skills or areas of specialization associated with each general expertise field (figure 5).



**Figure 5: Expertise**

#### **D. Directorate Search**

This option gives the user the ability to search for experts based on their departmental location. The first drop-down menu helps the user select a directorate, and the second drop-down menu allows the user to choose a particular branch or office within that directorate, allowing for more specific results. The directorate database was constructed at the KM Lab using information available through the GSFC's directorate web pages (Figure 6).

#### **E. Advanced Search**

This option will allow a more refined search for experts using a combination of: Name, Directorate or Office, or Career Summaries. There are three "combo" boxes, one for each field. The user can type a word in one or more of these fields, yielding results based on that information (Figure 7).

Search Expert Seeker - Directorate - Microsoft Internet Explorer

**NASA expertseeker**  
online knowledge repository of experts

**goddard space center renaissance**

please select a search method

name  
expertise  
directorate  
advanced  
web page

Directorate:  
300.0 - Flight Assurance

Offices:  
-Select All-  
-Select All-  
301.0 - System Review Office  
302.0 - Systems Safety and Reliability Office  
303.0 - Assurance Management Office

Submit Main Page

Figure 6: Directorate

Search Expert Seeker - General Search (Keywords) - Microsoft Internet Explorer

**NASA expertseeker**  
online knowledge repository of experts

**goddard space center renaissance**

please select a search method

name  
expertise  
directorate  
advanced  
web page

Search by Selecting more than One Field

Name: Naus  
Directorate / Office: Applied Engineering

Career Summaries  
Information Technology

Submit Reset Main Page

Figure 7: Advanced Search

## F. Web Page Search

This option accesses the Web Page Search Page. The user types the keywords in the blank box, and Expert Seeker will search through the employee web pages for a match. The results will display a list of experts whose web pages that contain the keyword being researched. This search differs from the expertise web-based search in that the user can search based on any keyword and not necessarily and expertise area.

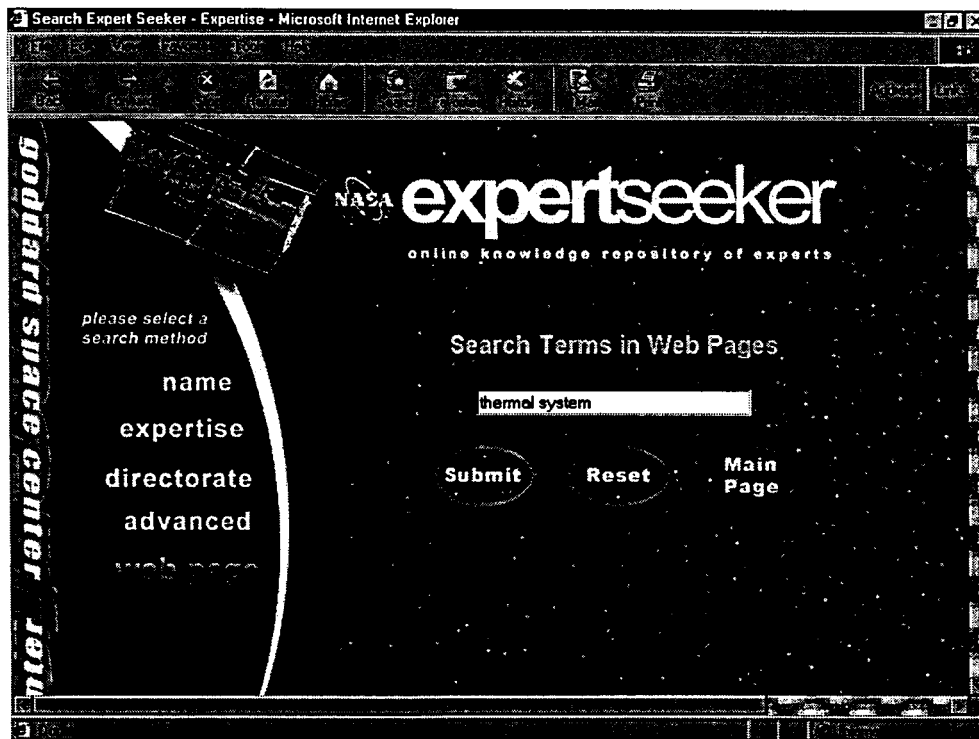


Figure 8: Web Page Search

## VII. Operators

Expert Seeker users can make the most of this search option with the help of "operators". Operators are commands used to specify more than one search word or search element. The operator tells Expert Seeker whether the user wants all the words to be present in the document to count as a match, or if the user wants any of the elements in the document to be retrieved.

The User can use more than one operator in a query. Most operators require the placing of angle brackets (< >) around the operator to clearly distinguish its meaning. However, Expert Seeker does not require the use of angle brackets for default operators and modifiers; <AND> and <OR> are assumed to be default operators and <NOT> is assumed to be a modifier when used.

There are various types of operators:

### **A. Concept Operators**

Concept Operators can be defined as word combination operators that allow the user to search for combinations of words, phrases, or a word and a phrase. The concept operators are:

1. <AND> --The "and" operator determines that all words in the search request must be considered for a match.
2. <OR> -- The "or" operator determines that any one or more of the words can be considered a match.

### **B. Evidence Operators**

Evidence operators can be defined as word operators that allow the user to search for words that are slightly different from the search words the user actually specified. These operators apply to the single word that immediately follows it. To use an evidence operator, the user must first enter the operator in angle brackets and then the word. The evidence operators are:

1. <STEM>. -- The "stem" operator finds all standard grammatical variations of the word specified by the user. In other words the system takes each word and utilizes its root to retrieve results containing common terms with the same stem.
2. <WILDCARD>. -- The "wildcard" operator finds all words with the string, including anything before or after the asterisk "\*" or question mark "?". The characters "\*" and "?" are automatically assumed to be wildcard characters, even if they are not specified to be operators.
3. <WORD>. --The "word" operator finds the exact spelling with no variations.
4. <SOUNDEX>. -- The "soundex" operator finds all the words that sound like the word entered by the user.
5. <TYPO>. -- The "typo" operator finds all the words that are spelled similarly.

### **C. Proximity Operators**

The Proximity operators are used to specify how close together search words must be to each other within a document in order for that document to count as a match. To use this type of operators, the user must first enter a key word, then the operator in angle brackets, and then another key word. The proximity operators are also activated if the user first enters the operator in angle brackets immediately followed by the key words in parenthesis and separated by commas. The Proximity operators are:

1. <SENTENCE>. -- The "sentence" operator specifies that the user is interested in documents where the key words are found in any order within a sentence in order to retrieve it as a match.
2. <PARAGRAPH>. -- The "paragraph" operator specifies that the user is interested in documents where the key words are found in any order within a paragraph in order to retrieve it as a match.

3. <NEAR/N>. -- The "near/n" operator specifies that the user is interested in files where the key words are found in any order within an "n" number of words apart from each other in order to be considered a match.
4. <NEAR>. -- The "near" operator specifies that the user is interested in documents where the key words are found in the same manuscript in any order; the closer the proximity, the higher the score.

## VIII. Challenges

One of the challenges in developing Expert Seeker was acquiring access to the MARS database in a timely manner. The largest obstacle was that at the KM Lab we do not support Sybase databases, since the underlying server technology that supports the KM Lab is SQL Server 7.0. Furthermore, it was also difficult obtaining access to the X.500 but this problem was solved when Mr. Pierce sent us Microsoft Access and Microsoft Excel files containing both the MARS and X.500 databases via FTP.

Another challenge involved the migration of the MARS database from its original format in Microsoft Access to create a relational database (SQL Server 7.0). This required the generation of several scripts in SQL Server 7.0 which would allow us to relate the information in the MARS database to the corresponding data in the X.500 database which was already in Expert Seeker's database. Even more challenging was the fact that not everyone in the X.500 database had their skills logged into the MARS database. Steve Naus later explained that the skill data from MARS only contains skills for about half of the GSFC employees, only those employed by Code 500 and Code 700. Unfortunately, the remainder of the core competency skills from other organizations are not available, or may be out-of-date, since the data was collected in 1997 and has not been maintained.

By far the largest barrier was obtaining access to the GSFC Intranet. After the meeting in Miami, Nancy Laubenthal followed up with Jerome Bennett regarding the NASA Center for Computational Sciences (NCCS) account, which we had requested in May of 2000. To activate the account, we filled out a User Agreement Form, and sent it on June 06, 2000. A couple of days later we received an email from Tim Burch, on behalf of the NCCS, welcoming us to the NCCS User Community. At this time we should have had access with our NCCS user ID. However, we tried logging-in via Telnet but we were unsuccessful. We called Tim Burch at the NCCS, but he was unavailable at the time, so Pravin Gohel assisted us, with no luck. We also e-mailed Daniel Russ from TAG for support and he suggested that we used Secure Shell (SSH) instead of using Telnet. We downloaded a 30-day trial version of SSH, but this did not work either. When we tried to log in we got a message specifying that the host, in this case *charney.gsfc.nasa.gov* and *suomi.gsfc.nasa.gov*, used SSH1 and not SSH2, which is what we had. Dr. Susan Hoban also tried to find the problem by getting us in contact with Joel Sachs, who notified us where to find SSH1 for download. Dr. Susan Hoban found out that our account was on *Jaylee.gsfc.nasa.gov*. Although we used SSH1 at the location specified, we were still not able to log in. Sue Kaltenbach later informed us that we had

to provide the IP addresses from where we were going to access the system, Dan Russ also informed us that the NCCS had created accounts for us on an NCCS debian linux box *nccsx4.gsfc.nasa.gov* and that we would gain access to *nccsx4* using SSH1 from our computers with IP addresses 131.94.129.[132,133,134,135]. Despite all the help that we received from numerous NASA-GSFC personnel including Jerome Bennett, Nancy Laubenthal, Tim Burch, Pravin Gohel, Dr. Daniel Russ, Dr. Susan Hoban, Joel Sachs, Sue Kaltenbauch, and Mike McGunigale, it was such a difficult challenge that we were never able to log in to NCCS.

## **IX. Immediate Objectives**

We are currently finishing the documentation for Expert Seeker. We are preparing two different manuals:

- **User's Manual:** A guide for users to learn how to access and use Expert Seeker. It points out the advantages of the system, includes a description of the different search modes and how the results are displayed, as well as a search guide.
- **Reference Manual:** A guide that describes the code, databases and scripts. This manual is intended for programmers responsible for the maintenance of the system. We are also prepared to make any changes suggested by the Advisory group during the final presentation.

## **X. The Technologies used to develop Expert Seeker**

- Cold Fusion 4.0, CGI, Python for Coding and Programming
- Photoshop 5.0, HTML and other Web Development tools for the User Interface
- SQL Server 7.0 for the Databases
- Verity '97 for Search Capabilities
- In-house codification of the Term Frequency-Inverse Document Frequency (TFIDF) and Information Retrieval (IR) Algorithm.

## **XI. Relevant publications related to this project**

Becerra-Fernandez, I., "The Role of Artificial Intelligence Technologies in the Implementation of People-Finder Knowledge Management Systems" in Proceedings of the 2000 American Association for Artificial Intelligence (AAAI) Spring Workshop "Bringing Knowledge to Business Processes", March 2000, Menlo Park, California.

Becerra-Fernandez, I., "Knowledge Management Systems", in Proceedings of the Performance Computing and Communications Program/Computational Aerosciences (HPCC/CAS) 2000 Workshop, February 2000, NASA Ames Research Center, Moffet Field, California.

Becerra-Fernandez, I. "Facilitating the Online Search of Experts at NASA using Expert Seeker People-Finder" In Proceedings of the Third International Conference on

Practical Aspects of Knowledge Management to be held on October 30-31, 2000, Basel,  
Switzerland

## **XII. Attachments**

# The Role of Artificial Intelligence Technologies in the Implementation of People-Finder Knowledge Management Systems

Irma Becerra-Fernandez, Ph.D.

Assistant Professor

Decision Sciences and Information Systems  
Florida International University, Miami, FL 33199  
Tel: 305-348-3476, E-mail: [becferi@fiu.edu](mailto:becferi@fiu.edu)

## Abstract

The development of Knowledge Management Systems (KMS) demands that knowledge be obtained, shared, and regulated by individuals and knowledge-sharing organizational systems such as Knowledge Repositories. One kind of Knowledge Repository, known as Knowledge Yellow Pages or People-Finder Systems, are repositories that attempt to manage knowledge by pointing to experts possessing specific knowledge within an organization. This paper presents the insights, challenges and future plans for the development of two People-Finder KMS: the Searchable Answer Generating Environment (SAGE), and the Expert Seeker. Here we also discuss the role that Artificial Intelligence technologies play in the development of People-Finder KMS and in automating the profile-maintenance.

## Introduction to Knowledge Management Systems

Knowledge Management Systems (KMS) have been defined as "an emerging line of systems [which] target professional and managerial activities by focusing on creating, gathering, organizing, and disseminating an organization's 'knowledge' as opposed to 'information' or 'data'" (Alavi and Leidner 1999). It has been observed that KMS currently underway at most organizations fall into three categories (Becerra-Fernandez 1999a):

1. Educational KMS: To elicit and catalog tacit knowledge, and at the same time serve as an educational tool.
2. Problem-Solving KMS: Organizations with significant intellectual capital require eliciting and capturing knowledge for reuse in solving new problems as well as recurring old problems.
3. Knowledge Repositories: The majority of the KMS in place. One kind of Knowledge Repository is known as Knowledge Yellow Pages or People-Finder Systems, are repositories that attempt to manage knowledge by holding pointers to experts who possess specific knowledge within an organization.

The paper presents insights from the development of two examples of such People-Finder KMS: the Searchable Answer Generating Environment (SAGE), and the Expert Seeker. This paper discusses insights and lessons learned from the development of these two systems. Finally, it presents the role of technology in automating the process of profile-maintenance, as well as future plans for the integration of Artificial Intelligence technologies in the

development of People-Finder KMS.

## The Searchable Answer Generating Environment (SAGE) KMS

The NASA/Florida Minority Institution Entrepreneurial Partnership (FMIEP) grant is funding the development of the Searchable Answer Generated Environment (SAGE), which is in the category of People-Finder KMS (Becerra-Fernandez 1999b). The purpose of this KM System is to create a repository of experts in the State of Florida (FL) State University System (SUS). Previous studies have pointed out that there is a void in the ability to identify the capabilities in the FL SUS (Kotmour 1998). Currently, each State University in Florida keeps a database of funded research, but these databases are disparate and dissimilar. The SAGE KM System creates a single repository by incorporating a distributed database scheme, which can be searched by a variety of fields, including research topic, investigator name, funding agency or university. As NASA-Kennedy Space Center (KSC) looks to develop new technologies necessary for the continuation of their space exploration missions, their need to partner with Florida SUS experts becomes evident.

The main interfaces developed on the query engine use text fields to search the processed data for key words, fields of expertise, names, or other applicable search fields. The application processes the end user's query and returns the pertinent information. The purpose of the SAGE KMS is to unify myriad data collections into one database collection that could easily be mined for relevant data. The benefits of SAGE are:

1. SAGE is a repository of Intellectual Capital within the state of FL SUS.
2. SAGE helps locate FL SUS researchers for collaboration with industry and federal agencies, thus increasing the potential for research funding to the SUS.
3. SAGE enhances communication and allows more visibility for FL SUS experts, making universities more marketable.
4. SAGE combines and unifies existing data from multiple sources into one user web-accessible interface.

The SAGE system addresses an important KM problem:



giving a user access to distributed knowledge, through a web-based Graphical User Interface.

### **The Technologies to Implement SAGE**

The development of SAGE was marked by two design requirements: the need to validate the data used to identify the experts, and at the same time minimize the impact of each of the universities' offices of sponsored research, who collect most of the required data. For this reason, we opted for taking the data structure in its native form and making necessary data cleansing at the SAGE server site. SAGE's strength rests in the fact that is built upon a criterion that is recognized as a valid indicator of expertise, actual funded-research grants received.

Although a number of database systems exist on the world-wide-web, which claim to help you find people with a defined profile, most of these tools rely on people to self-assess their skill against a predefined set of keywords. Self-assessment is inherently unreliable, and the results could be biased and hard to normalize. On the other hand, while a number of search engines are available on the web, the entity seeking for an expert has to use a combination of different tools in order to get find the appropriate information. With SAGE, all the information is easily accessible due to the versatility of its searching options, which allow you to refine the search until you get the degree of accuracy required.

SAGE is built upon the integration of the following technologies

1. Cold Fusion™ – An off-the-shelf Rapid and Integrated Development Environment.
2. Open Database Connectivity (ODBC) – allows middle-ware to interface with the database.
3. Verity's Search 97 – used to perform the Keyword search. It also allows the use of logic operators, which enhances the power of the search engine.

SAGE is online since August 16, 1999 at <http://sage.fiu.edu>.

One of the technical challenges faced during the design and implementation of this project was the fact that the source databases of funded research from the various universities were dissimilar in design and file format. Manipulating the data included the process of cleansing the data, followed by the data transformation into the relational model, and ultimately the databases migration to a consistent format (in this case SQL Server 7.0). One of the most important research contributions of SAGE is the merging of inter-organizational database systems through the use of correspondence tables, which function much like array pointers, and allow compliance to differing database formats. Future developments for SAGE include the development of algorithms that will facilitate the maintenance of SAGE.

### **The Expert Seeker KMS**

The NASA Faculty Awards for Research (FAR) is funding the development of Expert Seeker, which is in the category of People-Finder KMS. Previous Knowledge Management studies at KSC affirm the need for a center wide repository, which will provide KSC with Intranet-based access to experts with specific backgrounds. Currently KSC is reorganizing from an operations center into a research and development center. Expert Seeker aims to help locate intellectual capital within NASA-KSC, and is this particular characteristic what differentiates Expert Seeker from SAGE (the latter a KMS to find experts within the Florida universities). Expert Seeker will be used to search for experts located at KSC, although its use is expected to expand to other NASA Centers. The Expert Seeker KMS will be accessed via KSC's Intranet, in contrast the SAGE KMS which is on the world-wide-web is accessible through the Internet. Another important difference between SAGE and Expert Seeker is that the latter will enable the user to search for much more detailed information regarding the experts' achievements, including information such as intellectual property, skills and competencies, as well as the proeficiency level for each of the skills and competencies. The Expert Seeker KMS will provide access to competencies available within the organization, including items that are not typically captured by the typical Human Resource applications, such as completed past projects, patents, hobbies, and other relevant knowledge. This People-Finder KMS will be especially useful when organizing cross-functional teams.

The main interfaces on the query engine in Expert Seeker will use text fields to search the proposed data for keywords, fields of expertise, names or other applicable search fields. Expert Seeker will allow KSC experts more visibility, and at the same time allow interested parties to identify available expertise within KSC.

### **The Technologies to Implement Expert Seeker**

The development of Expert Seeker requires the utilization of existing data as much as possible. Expert Seeker will use the data in existing Human Resources databases for information such as employee's formal educational background, the X.500 Directory for the employee point-of-contact information, a Skills Database which profiles each employee's competency areas, and GPES, an employee performance evaluation system.

Information regarding skills and competencies, as well as proficiency levels for the skills and competencies needs to be collected, to a large extent, through self-assessment. Recognizing that there are significant shortcomings of self-assessment, we propose to use an increased reliance in technology to update employees' profiles, and thus place less reliance on self-assessed data. For example, we're proposing the use of Global Performance Evaluation System (GPES), an in-house performance

evaluation tool, to mine employees' accomplishments and automatically update their profiles. Typically, employees find it difficult to make time to keep their resumes updated. Performance evaluations, on the other hand, are without a doubt, part of everybody's job. We therefore seek to use this tool, augmented with appropriate queries, to inconspicuously keep the employees profiles up-to-date. Finally, a data mining effort of the document repository will also contribute to update employees' profiles. Based on the assumption that authors of documents in the repository are subject matter experts, therefore, mining the electronic document repository will contribute to keeping employees' profiles up-to-date in an unobtrusive way.

### The Role of AI in People-Finder KMS

Future developments for People-Finder systems such as SAGE and Expert Seeker include the development and integration of artificial intelligence (AI) technologies to enhance the capabilities of these systems. For example, data mining could enhance the process of updating profiles by mining the authors of documents in an electronic repository and identifying a correspondence with the topic of the document. Authors of documents in an electronic repository are experts in those knowledge areas; therefore, the profile of the contributors to the repository could be automatically updated with keywords related to the subject matter contribution. This data mining effort would result in a diminished reliance on self-assessment.

Furthermore, a data mining effort could be instrumental in clustering similar data objects together. For example, the data in SAGE is organized by grant awards and indexed by the Principal Investigator (PI) field. Through the use of a clustering tool (Mehrotra, Alvarado, and Wainwright 1999), data can be grouped into clusters of expertise, to reveal expertise areas that may not be currently defined. The implementation of the clustering technology will create a domain dictionary that will serve to increase the semantic domain of the keyword. In this fashion, relationships that may not be necessarily obvious may be identified also – a sort of “fuzzy matching.” The resulting “pseudo-keywords” may be saved for future re-use.

Another application of this clustering notion is the development of a “super” concept, which would allow to group experts together, developing a group-level of expertise. In the case of SAGE, grouping of researchers with completing areas of research from universities in the Florida State University System would result in virtual “centers of excellence”. This effort could reveal areas of strength that could otherwise go unnoticed in the organization. Additional developments in this area will be instrumental in the development of organizational training programs, designed to address the gap between what “is

known” and what “needs-to-be-known”.

In conclusion, our vision of People-Finder KMS fits well with the work to develop systems that seek to create an IT-support environment for knowledge workers. This is done through the use of intelligent assistants in a business process environment; keeping in mind that “an IT tool may only act as a facilitator for sharing, creating or retrieving knowledge, but never as a key player in creating, evaluating or contributing knowledge” (Schurr, Sttab, and Studer 1999).

### Acknowledgements

The author wishes to acknowledge NASA-KSC and Bethune-Cookman College, for financial support in the development of SAGE, under the auspices of the “Florida Minority Institution Entrepreneurial Partnership Grant”, grant number NAGO-0220. Thanks also to NASA-Headquarters for financial support in the development of Expert Seeker, under the Faculty Awards for Research (FAR-99), grant number NAG10-0259. The author also wishes to acknowledge James Jennings, Shannon Roberts, and Gregg Buckingham, and all of NASA-KSC who championed these initiatives. Finally, the author wishes to acknowledge the contributions of the students who work in the FIU Knowledge Management Lab and who collaborated in this research.

### References

- Alavi, M., Leidner, D. 1999. Knowledge Management Systems: Issues, challenges, and benefits. *Communications of the Association for Information Systems*, 1, article 7.
- Becerra-Fernandez, I. 1999a. This is rocket science: The Knowledge Repository of Kennedy Space Center. In *Proceedings of the Delphi's International Knowledge Management Summit (IKMS'99)*, San Diego, California.
- Becerra-Fernandez, I. 1999b. Searchable Answer Generating Environment (SAGE): A Knowledge Management System for Searching for Experts in Florida. In *Proceedings of the 12<sup>th</sup> Annual International Florida Artificial Intelligence Research Symposium*, Orlando, FL.
- Kotmour, T. 1998. Presentation at the Partners in Education Conference, Cocoa Beach, Florida.
- Mehrotra, M., Alvarado, S., and Wainwright, W. 1999. Laying a Foundation for Software Engineering of Knowledge Bases in Spacecraft Ground Systems. In *Proceedings of the 12<sup>th</sup> Annual International Florida Artificial Intelligence Research Symposium*, Orlando, FL.
- Schurr H., Sttab S., and Studer R. Ontology-based Process Support. 1999. In *Proceedings of the AAAI Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning*, Orlando, FL.

## **KNOWLEDGE MANAGEMENT SYSTEMS**

**Irma Becerra-Fernandez, Ph.D.  
Florida International University  
Department of Decision Sciences and Information Systems  
(305) 348-3476, becferi@fiu.edu**

The development of knowledge management systems (KMS) demands that knowledge be obtained, produced, shared, regulated, and leveraged by a steady conglomeration of individuals, processes, information technology applications, and a knowledge-sharing organizational culture. It has been observed that KMS currently underway at most organizations fall into three categories: educational, problem-solving systems, and knowledge repositories - which constitute the majority of the KMS in place. Educational KMS are used to elicit and catalog tacit knowledge, and simultaneously serve as educational tools. Problem-solving KMS are used by organizations with significant intellectual capital that require eliciting and capturing knowledge for reuse in order to solve new problems as well as recurring problems, based on experience gained from solving previous problems. Knowledge repositories, include repositories of organizational knowledge that exists in explicit form (e.g. system to store marketing-oriented documents), less structured databases of employees' insights and observations (e.g. "discussion databases" or "lessons-learned systems") and repositories that attempt to manage organizational knowledge by holding pointers to experts who possess specific knowledge within an organization. The latter category of KMS has been referred to in the literature as Knowledge Yellow Pages or People-Finder systems. This paper discusses the development of two examples of such people-finder KMS, the Searchable Answer Generating Environment (SAGE) and the Expert Seeker KMS. SAGE is a KMS used to identify experts in the Florida State University System (SUS). Currently, each Florida State University maintains information concerning funded research, but these databases are disparate and disjoint. The SAGE application creates one single web-enabled repository, which can be searched in a number of ways including Research Topic, Investigator Name, Funding Agency, or University. The Expert Seeker KMS, currently under development, seeks to help locate intellectual capital within KSC at all educational levels. The application will store the competencies available within the organization, including items that are typically not captured by Human Resources applications, for example, past projects that have been completed, patents, and other relevant knowledge. This repository will be especially useful when organizing cross-functional teams. This application combines and unifies existing data from multiple sources into one user accessible interface. Expert Seeker allows the identification of a researcher's expertise within a discipline and facilitates communication or a point of contact. Insights and lessons-learned gained from the development of these two systems are discussed. The process of profile-generation, and maintenance is also discussed. Finally, the role of technology in automating the process of profile-maintenance in order to diminish the impact that self-assessment introduces in the profile generation, as well as future plans are presented. The paper will be presented at the "AAAI Spring Workshop on Bringing Knowledge to Business Process", March 2000, and included in those proceedings. An on-line version will be available after March 20 at <http://www.fiu.edu/~cikm>.



# Facilitating the Online Search of Experts at NASA using Expert Seeker People-Finder

Irma Becerra-Fernandez, Ph.D.  
Assistant Professor  
Decision Sciences and Information Systems  
Florida International University, Miami, FL 33199  
Tel: 305-348-3476, E-mail: [becferi@fiu.edu](mailto:becferi@fiu.edu)

## Abstract

People-Finder Systems are knowledge repositories that attempt to manage knowledge by holding pointers to experts who possess specific knowledge within an organization. This paper presents insights from the development of Expert Seeker, an organizational People-Finder KMS that will be used to locate experts at the National Aeronautics and Space Administration (NASA). This paper discusses insights and lessons learned from the development of this system, and the role of technology in automating the maintenance of the expert's profiles. Expert Seeker represents an important first step towards achieving our objective of automatically and intuitively discovering and identifying intellectual capital within the organization. While several systems in place today rely on self-assessment, we look at the potential of artificial intelligence (AI) technologies, in particular, data mining and clustering techniques, to uncover and map organizational expertise.

## 1 Introduction to Knowledge Management Systems

Knowledge Management Systems (KMS) have been defined as "an emerging line of systems [which] target professional and managerial activities by focusing on creating, gathering, organizing, and disseminating an organization's 'knowledge' as opposed to 'information' or 'data'" [Ala99]. KMS currently in use at most organizations, fall into three categories [Bec99A]:

The copyright of this paper belongs to Irma Becerra-Fernandez Ph.D. Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage.

Proc. of the Third Int. Conf. on Practical Aspects of Knowledge Management (PAKM2000)  
Basel, Switzerland, 30-31 Oct. 2000, (U. Reimer, ed.)

<http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-34/>

1. Educational KMS: To elicit and catalog tacit knowledge, and at the same time serve as an educational tool.

2. Problem-solving KMS: Organizations with significant intellectual capital require eliciting and capturing knowledge for reuse in solving new problems as well as recurring old problems.

3. Knowledge repositories: Under the auspices of KM, tools historically used for singular, unrelated purposes are integrated to address the corporate memory problem. One type of knowledge repository is People-Finder Systems, also known as Knowledge Yellow Pages.

People-Finder Systems are knowledge repositories that attempt to manage knowledge by holding pointers to experts who possess specific knowledge within an organization. Several organizations in different business categories have identified the need to develop systems to help locate intellectual capital, or People-Finder KMS. The intent in developing these systems is to catalog knowledge competencies, including information not typically captured by Human Resources systems, in a way that could later be queried across the organization. A literary review and a table comparing the characteristics of hallmark People-Finder KMS in use in organizations today appears in [Bec00].

The paper presents insights from the development of Expert Seeker, an organizational People-Finder KMS that will be used to locate experts at the National Aeronautics and Space Administration (NASA). This paper discusses insights and lessons learned from the development of this system, and the role of technology in automating the maintenance of the expert's profiles.

## 2 Motivation: Developing an Organizational Knowledge Management Strategy

In order to assess the areas of Intellectual Capital for Kennedy Space Center, a Knowledge Management Assessment (KMA) was designed and implemented between the months of February and April 1998. The goals of this effort were:

1. To analyze the current types, sources and uses of knowledge in the organization;
2. To develop a detailed set of system specifications and implementation plan for future related activities;
3. To create a detailed plan for dealing with future needs; and
4. To gather data for the implementation of a prototype that will address some of KSC's Knowledge Management needs.

For this purpose, a series of assessment interviews were designed and implemented, with the cooperation of representatives of the majority of the functional groups at KSC. The goals of the interviews were to assist KSC in identifying key competencies and analyze the current knowledge architecture for the center. This step would ensure that the appropriate methodology is recommended at the end of this phase.

Following is a summary of the findings from the Knowledge Management Assessment of KSC's Knowledge Management (KM) needs and possible enhancements to the current KM Environment:

1. Expert Knowledge Elicitation and Virtual Mentoring Tools: Six of the eight interviewed groups identified the need for the implementation of tools to elicit, capture, and transfer the knowledge of experts acquired through years of experience at NASA.
2. An "Expert Seeker" Knowledge Management System: Six of the eight interviewed groups expressed the need for an application that holds pointers to experts with a particular background. This application would help locate Intellectual Capital within the center at all levels, from technicians to Ph.D.'s. The Expert Seeker application would store competencies available within the organization, including for all KSC employees completed past projects, patents, and their relevant expertise. An added benefit would be to include competencies outside NASA-KSC, for example those of subcontractors.
3. Collaborative Tools: Of the 8 technical groups interviewed, 6 expressed a need for Internet/Intranet based collaborative tools that capture knowledge as teams create it, integrated with an electronic document storage.
4. Decision Support and Expert Systems: The need for the implementation of KMS that would enhance decision making and would facilitate the decision process by incorporating knowledge factors from past projects that might prove useful and help make better, more educated decisions in the future.
5. Center-wide Lessons Learned Repository: Five of the eight interviewed groups expressed the need for a center-wide Lessons Learned Repository.

Following from the recommendations presented at the conclusion of this study, the KSC Executive team decided to fund the implementation of the KSC Expert Seeker People-Finder.

### 3 Summary of previous research in People-Finder KMS: The Searchable Answer Generating Environment (SAGE)

The NASA/Florida Minority Institution Entrepreneurial Partnership (FMIEP) Grant is funding the development of the Searchable Answer Generated Environment (SAGE) which is in the category of People-Finder KMS [Bec99B]. The purpose of this KM System is to create a repository of experts in the State of Florida (FL) State University System (SUS). Currently, each State University in Florida keeps a database of funded research, but these databases are disparate and dissimilar. The SAGE KM System creates a single repository by incorporating a distributed database scheme, which can be searched by a variety of fields, including research topic, investigator name, funding agency or university. As NASA-KSC looks to develop new technologies necessary for the continuation of their space exploration missions, their need to partner with Florida SUS experts becomes evident.

The SAGE system combines the unified database by masking multiple databases as if they were one. One advantage of this method is that there is no need to reconfigure the data to fit it into one template. This methodology provides flexibility to the users and the database administrator, regardless of the type of program used to collect the information at the source. Although the project SAGE is specific in nature, what was desired was to develop tools and techniques that would make managing these independent databases as seamless as possible. One of SAGE's advantages is that there is only one user point of entry at the web-enabled interface, allowing multiple occurrences of the interface and giving the end user deployment flexibility. The main interfaces developed on the query engine use text fields to search the processed data for key words, fields of expertise, names, or other applicable search fields. The application processes the end user's query and returns the pertinent information.

SAGE has been online since August 16, 1999 at <http://sage.fiu.edu>. Future developments for SAGE include such projects as the development of algorithms that will facilitate the maintenance of SAGE in a more automatic fashion. This inter-organizational system will require coding developments at both the SAGE server

and at each of the university's servers. A complete description of SAGE, including implementation details and results, appears in [Bec00].

#### **4 The Expert Seeker People-Finder System at Kennedy Space Center**

The NASA Faculty Awards for Research (FAR) is funding the development of Expert Seeker, which is in the category of People-Finder KMS. Previous Knowledge Management studies at KSC affirm the need for a center wide repository, which will provide KSC with Intranet-based access to experts with specific backgrounds. Currently KSC is reorganizing from an operations center into a research and development center. Expert Seeker aims to help locate intellectual capital within NASA-KSC, and is this particular characteristic that differentiates Expert Seeker from SAGE (the latter a KMS to find experts within the Florida universities). Expert Seeker will be used to search for experts located at KSC, although its use is expected to expand to other NASA Centers. The Expert Seeker KMS will be accessed via KSC's Intranet. In contrast, the SAGE KMS, which is on the world-wide-web, is accessible through the Internet. Another important difference between SAGE and Expert Seeker is that the latter will enable the user to search for much more detailed information regarding the experts' achievements, including information such as intellectual property, skills and competencies, as well as the proficiency level for each of the skills and competencies. The Expert Seeker KMS will provide access to competencies available within the organization, including items that are not typically captured by the typical Human Resource applications, such as completed past projects, patents, hobbies, and other relevant knowledge. This People-Finder KMS will be especially useful when organizing cross-functional teams.

The main interfaces on the query engine in Expert Seeker will use text fields to search the proposed data for keywords, fields of expertise, names or other applicable search fields. The application will process the end user's query and returns the pertinent information. The information will be collected from a conglomeration of multimedia databases, and the presented as queried. The purpose of the Expert Seeker KMS is to unify myriad data collections into web-enabled repository that could easily searched for relevant data. Prior to this project, there was no single point of entry into a unified repository that allowed identification of employees based on specific skills. Expert Seeker will allow KSC experts more visibility, and at the same time allow interested parties to identify available expertise within KSC. This

People-Finder KMS will help to identify a researcher's expertise, within a discipline, and to facilitate communication with a point of contact.

#### **5 Expert Seeker at Goddard Space Flight Center**

To further create synergies between the efforts to develop Expert Seeker at Kennedy Space Flight Center (KSC), a similar effort to prototype Expert Seeker at Goddard Space Flight Center (GSFC) was funded by the Center of Excellence in Space Data and Information Sciences. Efforts related to this proposal will attempt to mirror, as funds allow, some of the efforts currently underway at KSC, including:

1. System specification and selection of the organizational groups to prototype the GSFC Expert Seeker People-Finder.
2. Development of the GSFC knowledge taxonomy.
3. Design and development of the GSFC-Expert Seeker.
4. Implementation of the system prototype.
5. Testing of the system prototype.
6. Rollout.

It is expected that implementing the GSFC version of Expert Seeker will be to a large extent a replication of the ongoing efforts at KSC, in order to minimize duplication of efforts and maximize the return-on-investment for NASA. The resources that will be provided by this grant will serve to ensure generic features for this innovative system. Furthermore, implementation of Expert Seeker at GSFC will further validate the effectiveness of this KMS and ensure the development of a system that could potentially be of value to all of NASA. On the other hand, it is expected that the Knowledge Taxonomy for GSFC will differ from the one for KSC. But this requirement does not pose a concern, as Expert Seeker could be developed so the software could be "configured" with customizable knowledge taxonomy.

#### **6 The Technologies to Implement Expert Seeker**

The development of Expert Seeker is being accomplished with the use of the following technologies:

1. Cold Fusion 4.0, Java Script, Active Server Pages ASP (Coding and Programming)
2. SQL Server 7.0 (Databases)
3. Verity (Search capabilities)
4. Adobe Phototshop 5.0 (GUI)
5. HTML and other web development tools

The development of Expert Seeker requires the utilization of existing data as much as possible. Expert Seeker will use the data in existing Human Resources databases for information such as employee's formal educational background, the X.500 Directory for the employee point-of-contact information, a Skills Database which profiles each employee's competency areas, GPES, an employee performance evaluation system, and PRMS a project resource management system. Figure 1 depicts the architecture of Expert Seeker. Furthermore, other related information deemed important in the generation of an expert profile which is not currently stored in an in-house database system can be user-supplied, such as employee's picture, project participation data, hobbies, and volunteer or civic activities. Information regarding skills and competencies, as well as proficiency levels for the skills and competencies needs to be collected, to a large extent, through self-assessment. Recognizing that there are significant shortcomings of self-assessment, we propose to use an increased reliance in technology to update employees' profiles, and thus place less reliance on self-assessed data. For example, we are proposing the use of Global Performance Evaluation System (GPES), an in-house performance evaluation tool, to mine employees' accomplishments and automatically update their profiles. Typically, employees find it difficult to make time to keep their resumes updated. Performance evaluations, on the other hand, are without a doubt, part of everybody's job. We therefore seek to use this tool, augmented with appropriate queries, to inconspicuously keep the employees profiles up-to-date.

Finally, a data mining effort of the document repository will also contribute to update employees' profiles. Based on the assumption that authors of documents in the repository are subject matter experts, therefore, mining the electronic document repository will contribute to keeping employees' profiles up-to-date in an unobtrusive way. For this purpose, we are currently experimenting with the use of the Term Frequency Inverse Document Frequency (TFIDF) algorithm. The TFIDF algorithm is used as a measure of the uniqueness or relevance of a document within a collection of documents with respect to a specific keyword. TFIDF is calculated by the following formula:

$$w_{ij} = tf_{ij} \cdot \log(N/n)$$

where  $w$  is the TFIDF score for term  $i$  in document  $j$ ,  $tf$  is the frequency of term  $i$  in document  $j$ , and the inverse document frequency or  $idf$  is calculated by the logarithm of the total number of documents divided by the number of documents term  $i$  appears at least once. A term that

appears frequently in fewer documents will generate a higher TFIDF score for  $w_{ij}$  than a term that appears with comparatively high frequency but appears in many documents. Thus the TFIDF score is a measure of how relevant or unique a document is for a keyword in relation to a collection of documents. The resulting internal representation vector of the documents can then be searched by keyword. The TFIDF algorithm will be used within the Expert Seeker system to locate experts within the NASA Goddard Space Flight Center and NASA Kennedy Space Center by mining published documents within the Intranet of these organizations. This can be done periodically to keep the internal document representations up to date and to index new documents. The resulting TFIDF vector will be used for search queries. Documents that are returned as a query result will then be indexed by author name. The final result will rank authors according to those with the highest-ranking documents for that keyword and display these to the user as a subject-matter expert.

## 7 Challenges in the Implementation of People-Finder KMS

Previous research [Bec00] conducted to establish the parameters to design Expert Seeker application has demonstrated that one of the challenges in developing People-Finder KMS is related to the inherent shortcoming of self-assessment. Most of the People-Finder KMS in place today, except for example SAGE People-Finder or Mitre's Expert Finder [Kot98], rely on each employee to complete a self-assessment of competency, which is later used when searching for specific knowledge areas. The disadvantage of self-assessment is that the results of self-assessment are subjective, based on each person's self-perception, the results could be hard to normalize, and employees' speculation about its possible use could 'skew' the results. For example, one particular organization conducted a skills self-assessment study during a period of downsizing. This resulted in employees' exaggeration of their competencies, for fear they might have been laid-off. On the other hand, another organization made it clear the self-assessment would be used to contact people with specific competencies to answer related questions. This self-assessment caused employees to be overly modest about their skill profiles, for they would be required to put to test their specific knowledge. Furthermore, one People-Finder in place at Microsoft [Bec00] required supervisors to ratify their subordinates' self-perceptions, and assign a quantifiable value to it, a requirement that many organizations would find this requirement too taxing on their supervisors.



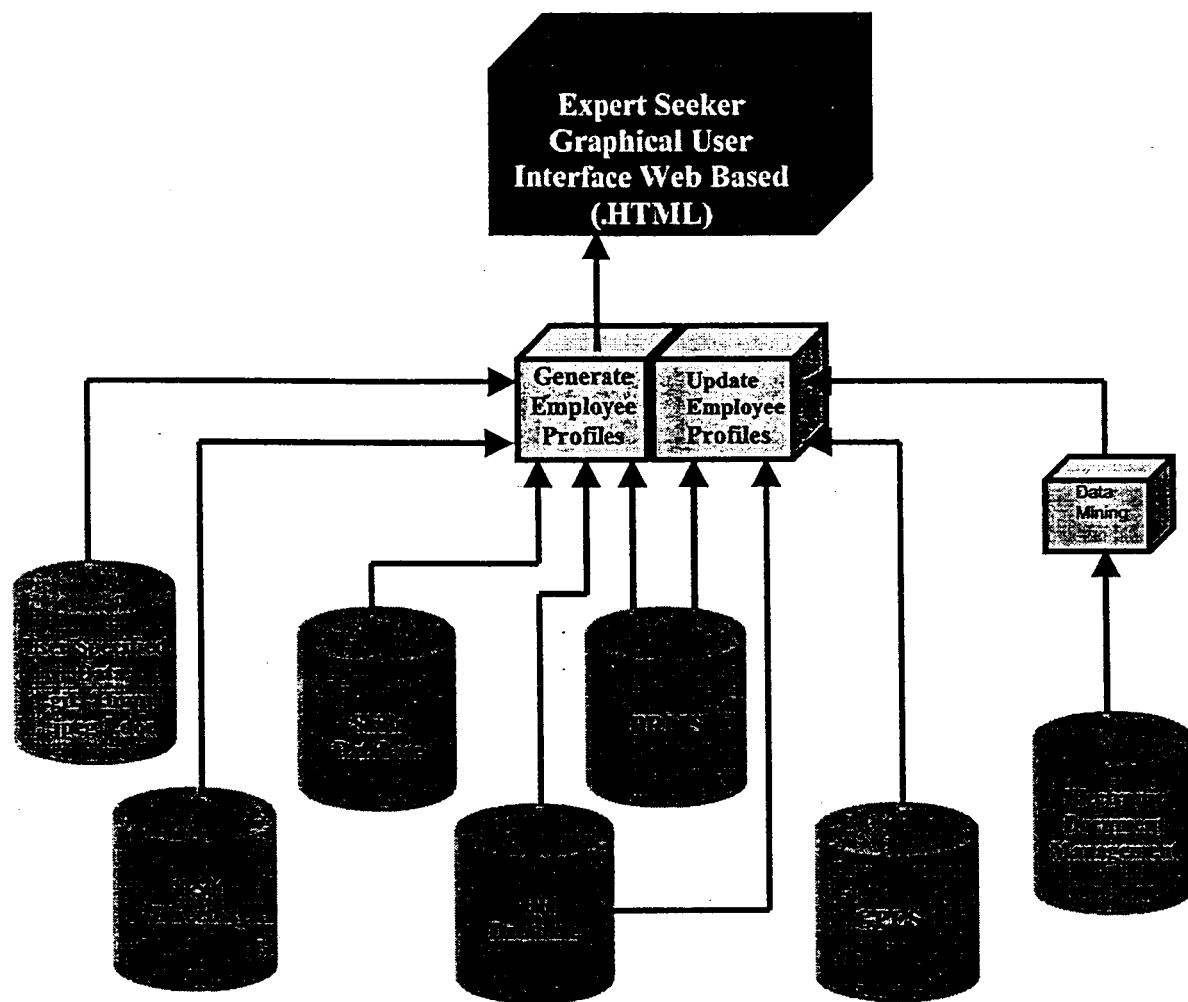


Figure 1:Expert Seeker Architecture

Another challenge in developing People-Finder KMS deals with the development of knowledge taxonomies. Taxonomy is the study of the general principles of scientific classification. Knowledge taxonomies allow organizing knowledge or competency areas in the organization.- In the case of People-Finder systems, the taxonomy is used to describe, and catalog people's knowledge, an important design consideration. Furthermore, knowledge taxonomies could be critical in the People-Finder system's success [Bec00]. People-finder KMS in place have addressed this consideration keeping in mind that:

1. Taxonomies should easily describe a knowledge area.
  2. Taxonomies should provide minimal descriptive text.
  3. Taxonomies should facilitate browsing, not complicate them.
  4. Taxonomies should have the appropriate level of granularity and abstraction. If the level is too high then it will be too complicated for the user, but if the level is too low it will not properly describe the knowledge areas.
- HP's CONNEX [Bec00] has been one of the successful systems in place that has developed a fairly functional taxonomy to describe people's knowledge. According to

Carrozza [e-mail 1999] HP's knowledge taxonomy is based on standards such as the U.S. Library of Congress Classifications (available online at <http://lcweb.loc.gov>) and the INSPEC Index (available online at <http://www.iee.org.uk/publish/inspec>), but is customized to their business area. Other firms have followed this model and have created their own taxonomies, as in the case for Microsoft's SpuD and for BA&H's Knowledge On-Line. According to Remeikis [phone interview, Sept. 3, 1999], this effort was successful. In contrast, the National Security Agency used a taxonomy based on a standard from the Department of Labor (O\*Net - available online at <http://www.doleta.gov/programs/onet>). Finally, according to Timothy Horst, Vice-President and Manager of Construction Resources and Technologies, Bechtel Construction Operations Incorporated also developed a taxonomy for their Knowledge Bank. It is based on standards developed by the National Center for Construction Education Research (NCCER - available online at <http://www.nccer.org>), but is only being used to catalog skills of manual workers [Phone interview, August 17 and 26, 1999]. While a number of work classification standards have been developed, that could be used to organize knowledge areas, we have not been able to apply any of these standards directly, without some thought and further development of the taxonomy. A deep analysis of the People-Finder KMS in place reveals that many attempts to create knowledge taxonomies are unsuccessful [Remeikis phone interview 1999] or sub-optimal [Carrozza phone interview and follow-up e-mail, 1999].

## 8 Practical Applicability of Expert Seeker

The Expert Seeker application has completed its first year development. The need for such a system to locate experts in an organization of more than 10,000 individuals, all with excellent qualifications, in order to reduce the time and effort spent in resolving issues pertaining to the R&D conducted at the different NASA centers. A preliminary usability study by NASA officials [Naus phone interview and follow-up e-mail, 2000] revealed minor weaknesses in the graphical-user interface that since have been corrected. Suggestions also focused on methods to access information, such as the capability of the system to allow combined searches. Expert Seeker, when completely implemented, is expected to become an effective tool in the management of knowledge required for new product and process development at NASA. Chris Carlson of NASA-Kennedy Space Center, envisions how Expert Seeker will be used in the future [Carlson phone interview and

follow-up email, 1999], as he describes a possible scenario:

*You are working in a project to build a new cryogenic handling storage facility. You encounter a problem, where upon testing, a valve fails. There is a design problem. You have two choices:*

- ◆ *The first choice is to go back through the same process with the same company and NASA engineers working the problem*

*or*

- ◆ *The second choice is to use Expert Seeker to organize the Rapid Answer Collaborative Knowledge Expert Team (RACKET). Using the expertise keyword 'cryogenics' Expert Seeker finds the following experts:*

1. *A collection of scientists from the University of Arizona for cryogenics studies;*
2. *A valve manufacturing expert from a plant in Detroit;*
3. *A cryogenic expert that worked on problems during shuttle that transferred to Marshall Space Flight Center.*
4. *In addition, the Expert Seeker uncovers a collection of technical white papers and lessons learned that NASA has published from similar projects.*

*The RACKET collaborates by video teleconference and the Internet to pinpoint the design problem, identify a feasible solution, and fixes the design problem in two days.*

## 9 Conclusions and Future Work

Results from the KSC Knowledge Management Assessment revealed the importance of a system to identify experts within the organization. Expert Seeker represents an important first step towards achieving our objective of automatically and intuitively discovering and identifying intellectual capital within the organization. While several systems in place today rely on self-assessment, we look at the potential of artificial intelligence (AI) technologies, in particular data mining and clustering techniques, to uncover and map organizational expertise.

Data mining technologies could contribute to updating employees' profiles. Based on the assumption that authors of documents in the repository are subject matter experts, mining the electronic document repository could contribute to keeping employees' profiles up-to-date in an unobtrusive way. Furthermore, clustering techniques

could be instrumental in clustering similar data objects together [Meh99]. In this case clusters of expertise, could reveal expertise areas that may not be currently defined. The use of clustering techniques provides the potential of creating a domain dictionary of "pseudo-keywords" that could serve to increase the semantic domain of the keywords, and which could be used to identify relationships that may not be necessarily obvious. Another application of this clustering notion is the development of a "super" concept, which would allow to group experts together, developing a group-level of expertise. Given the individual areas of expertise, these could be clustered together into groups of expertise or virtual "centers of excellence". In the case of Expert Seeker, grouping of experts within KSC or GSFC with complementing expertise areas could result in virtual "centers of excellence". This effort could reveal areas of strength that could otherwise go unnoticed in the organization.

#### Acknowledgements

The author wishes to acknowledge NASA-Kennedy Space Center and NASA-Headquarters under the Faculty Awards for Research (FAR-99), grant number NAG10-0259, as well as NASA-Goddard Space Flight Center and the CESDIS, contract number NAS5-32337 and subcontract number 5555-97-74, for financial support for the development of Expert Seeker. The author also wishes to acknowledge James Jennings, Shannon Roberts, Gregg Buckingham, Milt Halem, Susan Hubbard, Nancy Laubethal, and Jim Green, all of NASA who championed this initiative at KSC and GSFC. Finally, the author wishes to acknowledge the contributions of the students who work in the FIU Knowledge Management Lab and who collaborated in this research, specifically Claudia Alarcon, Greg Copeland, Hector Hartmann, Nelson Ospina, Haniel Pulido, Gabriela Rizzo, Juan Rodriguez, Heman Santiesteban, and Luis Felipe Villegas.

#### References

- [Ala99] Alavi, M., Leidner, D. *Knowledge Management Systems: Issues, challenges, and benefits*. Communications of the Association for Information Systems, 1, article 7 (1999).
- [Bec99A] Becerra-Fernandez, I. *This is rocket science: The Knowledge Repository of Kennedy Space Center*. Presentation at the Delphi's International Knowledge Management Summit (IKMS'99), San Diego, California (March, 1999).

[Bec99B] Becerra-Fernandez, I. *Searchable Answer Generating Environment . (SAGE): A Knowledge Management System for Searching for Experts in Florida*. Proceedings of the 12<sup>th</sup> Annual International Florida Artificial Intelligence Research Symposium (FLAIRS) - Knowledge Management Track, Orlando, Florida (May, 1999).

[Bec00] Becerra-Fernandez, I., *The Role of Artificial Intelligence Technologies in the Implementation of People-Finder Knowledge Management Systems*, Knowledge Based Systems special issue on Artificial Intelligence in Knowledge Management, forthcoming in 2000.

[Kot98] Kotmour, T. *Presentation at the Partners in Education Conference*, Cocoa Beach, Florida (October, 1998).

[Meh99] Mehrotra, M., Alvarado, S., and Wainwright, W. *Laying a Foundation for Software Engineering of Knowledge Bases in Spacecraft Ground Systems*. Proceedings of the 12<sup>th</sup> Annual International Florida Artificial Intelligence Research Symposium (FLAIRS), Orlando, Florida (May, 1999).